Duration as perceptual voicing cues in whisper

Yohann Meynadier, Sophie Dufour, Yulia Gaydina

Duration as perceptual voicing cues in whisper

Yohann Meynadier, Sophie Dufour & Yulia Gaydina

Université Aix-Marseille & CNRS UMR7309
Laboratoire Parole et Langage
Aix-en-Provence, France

In modal speech segmental durations are secondary phonetic marks of the voicing feature in many languages. The voiced obstruents show shorter durations than unvoiced ones; while the pre-consonant vowels are shorter before unvoiced consonants than before voiced ones[1].

In the one hand, auditory *(Kluender et 1988)*, articulatory *(Raphael 1975)* and/or main phonatory-aerodynamic *(Rothenberg 1968, Ohala 1983, 1997, 2011)* constraints could provide a phonetic ground of these differences in C or V duration relative to voicing. In the other hand, the durational cues of voicing are largely seen as linguistically controlled *(Lisker 1977, Ohala 2009)*. This can be supported by studies showing that perception of voicing is affected by the C and V durations[2-3] which enhance the possible phonological role of such phonetic details. But the particular share between physical and linguistic conditionings is still not well-known.

Because of the lack of voicing constraint, the **whisper could be a new paradigm** to asses the weight of physical *vs* phonological conditioning of segmental durations as function of voicing. No previous works had controlled C and V durations to assess the voicing perception in whisper, even if they showed that the voicing contrast is phonetically preserved in production and perception[4-7]. This study on French focuses on the **influence of voicing on segmental durations and of mismatched C and V durations on the perception of voicing in whispered speech**.

Production

Corpus

- Reading in **modal *vs* whispered** voice
- Lexical and non-sense isolated words
- 12 **voiced *vs* unvoiced obstruents**: p-b, t-d, k-g, f-v, s-z, ʃ-ʒ in median unstressed position
- 4 **French** speakers *(2 M & 2 F)*
- 5 repetitions

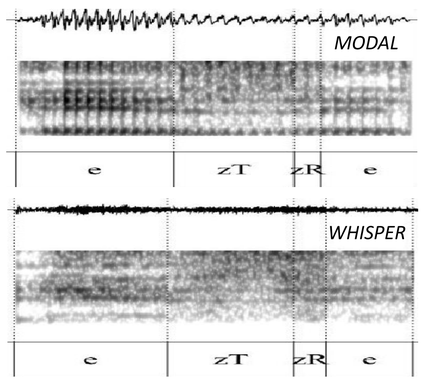
NON-SENSE WORDS
e**t**eve, e**k**eze, e**g**epe

LEXICAL WORDS
asyʁe / a**z**yʁe
de**b**ite / de**p**ite
e**k**ute / e**g**ute

Analysis

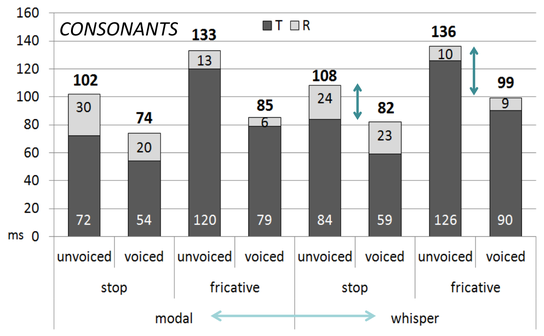
- **Acoustical durations** of **C** (steady-state *T*, offset *R*) and of the pre-consonantal **V**
- Repeated-measure Anovas on **duration** (ms)
- PHONATION: modal *vs* whisper
- VOICING: voiced *vs* unvoiced

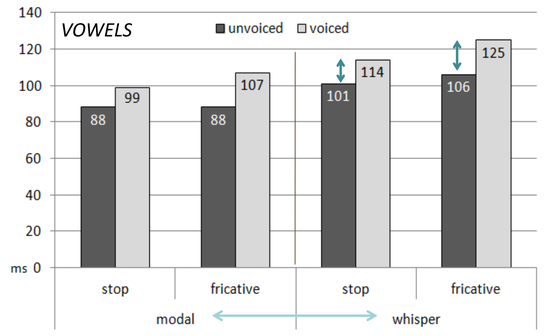
Non-sense and lexical words were pooled.



Results

PHONATION modal < whisper → slower rate





VOICING IN WHISPER (AS IN MODAL VOICE)

CONSONANTS
- $C_{[+vois]}$ < $C_{[-vois]}$
- $\Delta_{[vois]}$ remains for stops
- $\Delta_{[vois]}$ slighlty lower for fricatives

VOWELS
- before $C_{[-vois]}$ < before $C_{[+vois]}$
- $\Delta_{[vois]}$ remains for stops
- $\Delta_{[vois]}$ sligthly larger for fricatives

On modal speech, it confirms previous studies in French[1].

On whisper, it is in agreement with works on different languages[4] and with the only one study on French by Vercheran (2010).

The data show that the **phonological voicing contrast is also phonetically realized in whisper by segmental C and V durations**.

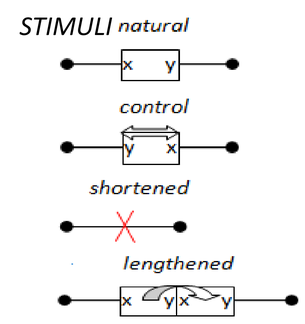
This study suggests that **duration differences as function of voicing are enough similar in modal and whispered speech**.

Perception

Experimental design

- Auditory stimuli : **non-sense minimal pairs**
- Binary visual choice : ' ona**b**é ' *vs* ' ona**p**é '
- 12 **voiced *vs* unvoiced obstruents** *(one M speaker)*
- **40 listeners** : 50% in **modal *vs*** 50% in **whisper**

Neutralized factors: balanced order, dB normalization



Stimuli

- **Natural** : no durational change, no acoustical modification
- **Control** : no durational change, signal inversion
- **Mismatch** : *from the empirical $\Delta_{[vois]}$ of observed durations in production*
  - $C_{[-vois]}$ shortened to $C_{[+vois]}$ duration (-30%)
  - $C_{[+vois]}$ lengthened to $C_{[-vois]}$ duration (+40%)
  - pre-$C_{[+vois]}$ V shortened to pre-$C_{[-vois]}$ duration (-15%)
  - pre-$C_{[-vois]}$ V lengthened to pre-$C_{[+vois]}$ duration (+15%)

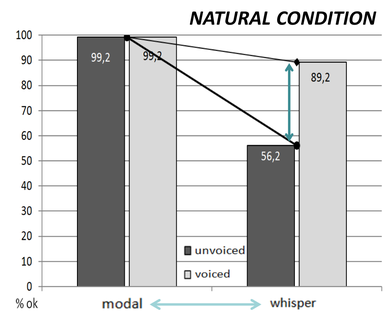
Analysis

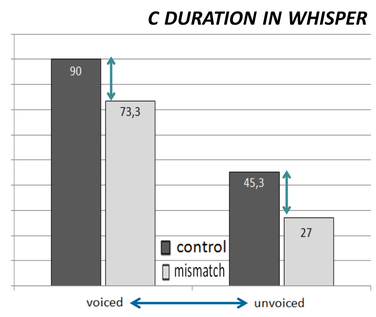
- Repeated-measure Anovas on **rate of correct responses** (%)
- PHONATION: modal *vs* whisper *(in natural condition)*
- VOICING: voiced *vs* unvoiced *(in control and mismatch conditions)*
- DURATION: control *(original duration)* *vs* mismatch *(opposite duration)*

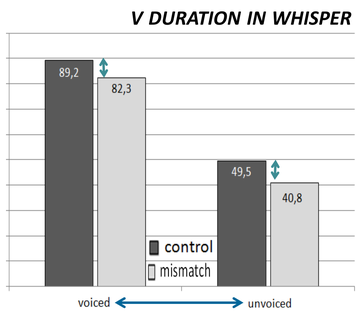
Stops and fricatives were pooled.

Results

PHONATION modal > whisper → lower recognition *(loss of intelligibility)*







VOICING

MODAL ≈ 100 % of correct recognition → **no effect**

WHISPER **$C_{[-vois]}$ < $C_{[+vois]}$**
- in all 3 conditions
- especially for non alveolar *(≠ t-d, s-z)*
- recognition of $C_{[-vois]}$ at the chance level

equivocal in littérature
(i) opposite direction[5]
(ii) variability accross speakers[6]
(iii) variability accross consonants[7] *as here.*

DURATION

MODAL → **no effect** of C[2] or V[3] duration
≠ previous works

WHISPER **mismatch duration < no change**

$C_{[-vois]}$ & $C_{[+vois]}$ recognition
- decrease for mismatch durations of C and V
- show more effect of C than V durations

$C_{[+vois]}$ recognition
- remains at high level: 90 % → 73-82 %
- no categorical switch of voicing perception

$C_{[-vois]}$ recognition
- remains very low: 45-50 % *(chance)* → 27-40 % *(no chance)*
- categorical switch to voiced percept

Two **unexpected results** are challenging for further works. First, the **unrecognized unvoiced C in whisper** is counter-intuitive: different investigations are needed (semantic priming?). Also, the **lack of duration effects in modal speech** could be due (i) to the uncontrolled spectral cues of voicing (laryngeal buzz, F1, noise intensity & freq.); *(ii)* to non-synthetic speech signal used here (≠ previous studies[8]).

However, the **C and V durations play a role in the perception of voicing in whisper**, i.e. even without any physical vibrations of the vocal folds.

Conclusions

Contrastive durations of C and V are produced as function of phonological voicing in French in modal and whispered speech.

Such **durational phonetic information play a role in voicing perception only in whisper and NOT IN MODAL VOICE**. This result confirms that listeners would use adaptative processes in the perception of phonological voicing.

As the whisper signal could be linkened to a spectrally altered speech, it meets studies showing that listeners switch from spectral to temporal cues in conditions of degraded or noisy speech (Winn et al. 2012).

Finaly, the **perception of whisper supports a possible linguistic ground of durational correlates of voicing**. But some laryngeal investigations of the voicing feature in whisper also suggest a potential aerodynamical conditionning[9].

References [1] for French: Chen 1970, O'Shaughnessy 1984, Bartkova & Sorin 1987, Lauefer 1992, Abdelli-Beruh 2004 – [2] Denes 1955, Lisker 1957, Wajskop & Sweerts 1973 (on French), Fledge & Hillenbrand 1986, Allen & Norwood 1988, Warren & Marslen 1989 – [3] Denes 1955, Raphael 1972, Hogan & Rozspypal 1980, Flege & Hillenbrand 1986, Allen & Norwood 1988, Crowther & Man 1992 – [4] Czech (Jovicic & Saric 2008), English (Sharf 1964, Parnel et al. 1977, Mills 2003, 2009, Kinsey 2005, Osfar 2011), Dutch (van der Velde & van Heuven 2011), Russian and Hungarian (Knyazev 1991), French (Vercherand 2010) – [5] Mills 2003, Vercherand 2010 – [6] Munro 1980, Higashikawa 1994 – [7] Fux 2012 – [8] Nittrouer 2004, 2005, Winn et al. 2012 – [9] Malécot & Peebles 1955, Weismer & Longstreth 1980, Higashikawa 1994, Mills 2009, Meynadier & Gaydina 2012, 2013.